**Challenge Selection Emails**

**General enquiry**

Thanks for the enquiry and for your interest in the Data Study Groups.

For more information about being a “challenge owner” please follow the link here: <https://www.turing.ac.uk/collaborate-turing/data-study-groups/get-involved-challenge-owner>

I would also suggest having a look at our current and past events as you will see the types of challenges that we have undertaken.: <https://www.turing.ac.uk/collaborate-turing/data-study-groups#previous-data-study-groups>

And finally, to get an idea of the types of outputs that you can expect from the event please take a look at some past reports here: <https://www.turing.ac.uk/collaborate-turing/data-study-groups#reports> .

If this all looks good to you, then please complete the challenge proposal form at the bottom of the challenge owner page. We can then make an initial assessment if your problem would be a good fit for the data study group and then we can respond to any specific queries.

**Proposal form confirmation of receipt:**

Thank you for submitting your enquiry form it has been received by the Data Study Group mailbox. The Data Study Group team will review your submission and be in contact within 30 days.

**Unsuitable challenge flat out rejection:**

Thank you for your interest in The Alan Turing Institute Data Study Groups.

The Data Study Group team has reviewed your challenge enquiry and assessed it against specific selection criteria which consider its readiness and suitability for the Data Study Group research model.

I regret to inform you that it is not considered to be a suitable challenge at this time.

You will appreciate that there is a high demand for Data Study Groups and we very selective in the Challenges which are progressed to ensure the best possible experience for Challenge Owners and for the students who are participating.

**Unsuitable due to data**

Your challenge has lots of potential that needs the right data to unlock it. This involves understanding how many instances of duplicate scans there are and whether there are sufficiently many, across sufficiently many unique products, and across sufficiently many steps in the process (factory, to consumer) in order to provide enough information to meaningfully address the challenge of identifying the difference between fraud and a legitimate cause. Also, there is not enough data across all factories to meaningfully address problems of outlier detection, or clustering (all the factories are far too different in output).

Perhaps once these areas are looked at we’ll be able to look at it for a September submission.

**Invitation for an initial call**

The DSG team have reviewed your Data Study Group proposal and think it has the potential to be a great challenge.

We’d like to set up a half hour call to discuss a couple of points in more detail.

Would the afternoon of Monday 3rd work for you?

**Unsuitable challenge response:**

The Data Study Group programme team has taken another look at your proposed challenge and feels that it isn’t yet ready to progress.

As you know, we are particularly concerned that DSG should be a learning experience for the students, as well as informative for the challenge owner.

The [two] enhancements that we would seek are:

* [Provision of an original data set, which means one that is not widely available and used, and one that contains a rich set of features that the participants can explore.
* Re-framing of the question from the classic “predicting markets” to one which focuses on a particular (practical or methodological) aspect that is truly novel, will be exciting and challenging to participants, and of appropriate scope for a five day DSG.]

We realise that this will be disappointing to you, but hope that you understand that this process of preparation and curation is what makes Turing’s DSG distinctive and beneficial for participants and challenge owners.

**Unsuitable challenge due to Crime relation:**

Thank you for your submission for a data study group challenge. Unfortunately this challenge is not something that we will be able to take forward.

There were 2 main concerns related to your challenge proposal that formed our decision.

1. In your description you mention a host of sources for potential data, not least some using protected personal data points. Even if anonymised, in combination with some of the other data sources you mention there would be a risk of being able to identify victims or other related individuals. At this time the Turing is not equipped to host such a collection of data sets in one repository for this kind of analysis.
2. In general, the quality of the data being provided would need to be very carefully scrutinised for accuracy (as you mention in the proposal). Any data in these areas are created under the influence of human biases (individuals with the app, police officers) or under the influence of particular policies (e.g. different degrees of stop and search, different priorities in the treatment of different crimes). Any data that comes from police from a time when a previous policy applied will no longer be accurate and cannot be used. This may also undermine the amount of data available and subsequently the accuracy of any ML models developed.

Given the time frame of the data study group, we feel that 1 week would not be enough time to fully explore this kind of challenge in the depth that would be required for such a topic. As it stands and given the constraints of the data study group, it would fail our ethic approval process for the reasons listed above.

Thank you again for the submission, it looked like an interesting challenge, and I am sorry we were not able to explore it further.

**Re-touching base**

I hope this email finds you both well.

I wanted to check in with you and see if you were still interested in hosting a Data Study Group challenge. We still have a few spaces remaining for our April addition or failing that the September or even December 2020 event.

We would still very much like to work with Predina so please do let us know if you would like to resume the process.

**CO pack and invitation to scope:**

Following your [call/meeting/talks with JM/CL ect] we would really like to organise the high-level scope of you challenge and think it has great potential as a Data Study Group Challenge.

This will take the form of a 1-hour consultation with one of our DSG science leads. Together you will iterate what exactly your challenge is hoping to address, briefly discuss your data and refine what is expected from you as a challenge owner and what you can expect from the Alan Turing Institute.

I have attached the Challenge Owner pack which explains the process in more detail. We have found this to be an essential resource so please do give it a read in full, feel free to contact me if you have any questions about anything in the pack.

Please let me know your availability so we can schedule the high-level scope of your challenge.